

# MODELING SPEED DOUBLING IN CARNATIC MUSIC

Srikumar K. Subramanian, Lonce Wyse, Kevin McGee

National University of Singapore

Department of Communications and New Media

{srikumar.k.subramanian, lonce.wyse, mckevin}@nus.edu.sg

## ABSTRACT

We consider the problem of modeling the feature relationships between multiple speed renditions of parts of compositions called *varnams* in Carnatic music, discuss related work in speech and singing synthesis and in synthesizing Carnatic music from solfège notation, present style dependent but arguably *rāgā* independent rules for simplifying and adjusting *gamakās* (continuous pitch movements) in the slower speed performance of one composition to derive the more rhythmic double speed performance and find that the performance derived using these rules compares favourably with the double speed rendition by the same artist.

## 1. BACKGROUND

Salient musical features that depend on speed occur in some musical genres such as jazz and in Indian classical music where performers are known to alter musical details of a composition to suit different speeds. In south Indian classical music (called “carnatic music”), the ornate *gamakās* (continuous pitch movements) used in a slower speed are simplified during performance at higher speeds and have a greater rhythmicity to them than their lower speed counterparts. The movements in the higher speed renditions are fewer and appear to follow a rhythmic pulse determined by the composition’s “*tālā*” (time structure). A transformation involving detail reduction while increasing such rhythmicity appears to be intricate and raises the question of how much genre knowledge is needed to execute it. The nature of these speed related transformations is the subject of this paper.

Within carnatic music, compositions in the category of *varnam* feature sections that are performed in multiple speeds within a single concert performance. *Varnams* therefore are suitable material for studying the changes that a performer makes to the slower speed rendition when performing it in a higher speed. A typical *varnam* consists of four parts, with the first three making up the first half of the composition. The first three parts (*pallavi*, *anupallavi* and *muktayisvaram*) are performed first at a slow speed and followed by one or more higher speed versions related to the original speed by simple integer factors. A *varnam* performance is guaranteed to at least feature a speed doubled version of the first three parts though a performer may choose additional speed multiples such as 3/2, 3, 4,

5, 6, 7, 8 and 9. The *tālā* (time structure or meter) is kept constant throughout the different speeds. The fourth part - *caramam* - is by convention performed at double the speed at which the piece is begun and in this case, the *tālā* is also doubled in speed. The *caramam* therefore provides additional raw material to study the characteristics of *gamakas* at higher speeds.

*Gamakās* that feature in the slower speeds cannot be preserved in the faster speeds. Since the slower speed already packs much more detail per note, a direct speed up would require absurd levels of detail in higher speeds that will be impossible for a performer to execute and will overload listeners. The work presented here is an attempt to model the kind of detail reduction that happens when increasing performance speed.

## 2. MOTIVATION

Our long term goal is to develop a synthesizer for the sparse and discrete “prescriptive notation” that is used for musical communication in the Carnatic genre.<sup>1</sup> Though the prescriptive notation omits the all-important *gamakās* — complex continuous pitch movements — that characterize the genre, trained musicians are able to fill in these details. Therefore a synthesizer for prescriptive notation can be said to capture the knowledge that a trained musician brings to the interpretation of a sparsely notated composition. Understanding the influence of speed on the choice and structure of *gamakās* is an important part of this larger synthesis problem. Performing such detail reduction of *gamakās* when given only the slower speed performance is also an important skill for a student of the genre. Therefore, computer modeling of this transformation, in addition to contributing to the musicology of the genre, may have pedagogical applications.

We now discuss work that has been done in the related areas of jazz swing modeling, text to speech synthesis, expressive music synthesis and *gamakā* synthesis in carnatic music.

<sup>1</sup>The term “prescriptive notation” was introduced by the ethnomusicologist Charles Seeger to denote notation forms that serve as instructions for performers, in contrast with “descriptive notation” which captures the details of a specific performance after the fact [13].

### 3. RELATED WORK

In considering musical features that depend on speed, jazz “swing” has been a subject of considerable study. In jazz, swing ratios - the ratio of the longer note or beat to the following shorter one - executed by performers are known to change with tempo [6, 4, 2, 3]. In [4], Friberg and Sundström present that swing ratios vary from 3.5:1 at slow tempi to 1:1 at fast tempi. In [6], Honing reports that professional jazz drummers have “enormous control over their timing” within a precision of milliseconds. Nevertheless, swing ratios are not kept constant and are “varied systematically with tempo”. Honing also notes that this is “in line with the more general hypothesis that expressive timing in music performance does not scale proportionately with tempo”.

Speech intonation models deal with the generation of the fundamental frequency contour - known as the “F0 contour” - are an important component of prosody and are related to *gamakās*. The most common model used for generating F0 contours for speech is the Fujisaki model which has been applied to both speech and singing [10]. According to this model, the F0 contour is generated as the response of a second order linear system to a sequence of discrete linguistic commands [5]. The “tilt intonation” model developed by Taylor and Black [19, 21] views the F0 contours of speech as a series of pitch “excursions” and describe each using an extent, a duration and a “tilt” parameter which varies from -1 (a pure fall) through 0 (a rise followed by a fall) to +1 (pure rise). Porte and Heuft’s “maximin-based description” uses another parameterization that is similar to Taylor’s model. They specify a contour by identifying F0 maxima, their times and their left and right slopes [11]. The minima are implicit in this model and sinusoidal interpolation of F0 is used to generate the complete contour using this information. Regarding the question of “naturalness” of such intonation models, Taylor notes in [20] that “the linguistic justification for any existing intonation systems are weak”, though Fujisaki does provide physiological justifications for his model. Modeling “speaking rate” control is another relevant area of speech synthesis where non-linear temporal stretching is used to preserve intelligibility of speech. Vowels and consonants, in particular, are time stretched by different amounts. For example, in [22], Yoshimura et al describe how to implement such speaking rate controls using hidden Markov models.

Expressive music synthesis systems also use speed dependent expression parameters in their models. Sundberg et al in [18] present musical expression rules connecting duration and pitch leaps in their attempts towards expressive synthesis of baroque music. Their system “Director Musicés” has been a long-standing top contender in the annual *RenCon* contest [8]. Berndtsson, in [1], identifies duration dependent expression parameters for singing synthesis such as “durational contrast”, “double duration” and “swell on long tones” that depend on the time available to execute these expressions. It is interesting to note

that both these systems for different forms of musical expression use a rule based system that was originally built for text to speech synthesis.

In the domain of carnatic music, M. Subramanian [17] has built an expert system featuring a per-raga database of context dependent rules for the automatic derivation of *gamakās* from the sparse prescriptive notation. Such a synthesizer is one way to approach the speed dependency problem, since it needs to handle both the normal speed and the double speed renditions, both of which have the same prescriptive notation. His system sits within the framework of his “svara notation”<sup>2</sup> synthesis program called “Gaayaka” [15]. In his approach to synthesizing prescriptive notation, the preceding and following pitches of a notated pitch are used as the melodic context, together with the duration of the pitch to be expanded to determine the final phrase.<sup>3</sup> To deal with the fact that different gamakas are needed for faster phrases, Subramanian populates the database with entries for five different duration ranges for each pitch triad. To account for multiple interpretations of a given notation fragment, Subramanian presents the possibilities as choice to the user of his Gaayaka system. Though this is a viable approach to handling the speed dependency of gamakas, it doesn’t account for the possibility that the detail reduction upon increase of speed might have a pattern to it, potentially spanning multiple *rāgās*.

In the following sections we describe this speed doubling problem and present a logical model for the speed dependencies of gamakas for one *rāgā*. Such a model can capture some of the musical understanding that a practicing musician might bring to the act of changing speeds, besides helping to reduce the complexity of phrase databases such as used by Subramanian.

### 4. PROBLEM

A qualitative assessment of the relationship between the “first speed” and “second speed” renditions of portions of a *varnam* indicates that the latter has a stronger rhythmicity to it and has fewer details in its melodic movements. The problem, therefore is to determine the extent to which the second speed rendition can be derived from the first speed rendition. The nature of this transformation appears discrete and therefore different from the continuous nature of the transformations in text to speech and expressive music synthesis discussed in §3.

The parts of the problem are -

1. Reducing *gamaka* details in the double speed relative to the slower speed rendition.
2. Modeling the relationship between *gamakās* and the underlying pulse derived from the tala.

<sup>2</sup>We use the term “prescriptive notation” for the same, ignoring the additional notational elements introduced in Gaayaka.

<sup>3</sup>The term “pitch” is inaccurate and the term “svara” which stands for solfege *and* pitch as a unified entity better captures what is found in the prescriptive notation. However, we stick to “pitch” because it is accessible to a wider audience.

3. Adhering to the *rāgā*'s restrictions on the types of *gamakās* that are allowed.
4. Calculating microtonal pitch adjustments for the double speed rendition based on perceptual criteria.

We now discuss our approach to this problem.

## 5. METHOD

We approached the modeling problem by selecting a performance of a *varnam*, transcribing the necessary gamaka details into a numeric representation verified by re-synthesis, analyzing the characteristics of multiple speeds present in the performance, using the resultant rules to transform the slower speed performance into a higher speed performance and finally comparing the generated performance with the real performance.

### 5.1. Material selection

The material selected for this study was the performance of the *varnam* titled “Karunimpa” in the *rāgā* “*Sahānā*” and *tālā* “*Ādi*” by the late *vīṇā* maestro Smt. Rajeswari Padmanabhan. The structure of this *varnam* is shown in Table 2 and is typical of most *varnams*. A *varnam* was chosen because its length and the variety in its melodic content, which enables us to work with its internal compositional consistencies.

*Sahānā*, whose scalar structure is shown in Table 1, has *gamakā* characteristics that set itself apart from other ragas with similar scalar structure. Despite its uniqueness, it is not considered a complex raga and can therefore help to shed light on gamaka characteristics that are likely to transcend *rāgās*.

*Ādi tālā* is an 8-count time cycle grouped as  $4 + 2 + 2$ . For the portion of the *varnam* we're studying, the *tālā* occurs in the doubled length form known as “2 *kalai*” where it is effectively a 16-count cycle grouped as  $8 + 4 + 4$ . Therefore, in this paper, we use the word “count” in the latter sense of 16 counts per cycle.

Ascent	<i>CDEFGFABbC</i>
Descent	<i>CBbAGFEFDEDC</i>

**Table 1.** The scalar structure of *rāgā* “*Sahānā*”.

### 5.2. Transcription by re-synthesis

The recording was transcribed into a numerical representation that captures the pitch and time aspects of gamakas to an adequate degree of detail for re-synthesis. Aural A/B comparison of individual “notes” in the performance against the re-synthesis was used to verify the transcription. A simple additive synthesis instrument built in SuperCollider was used for the re-synthesis and was adequate for the task at hand. This process yielded the minimal pitch-time data needed for further analysis.

Part-1	Pallavi Anupallavi Muktayisvaram
Part-1 (2x speed)	Pallavi Anupallavi Muktayisvaram
Part-2 (2x speed)	<i>Caranam</i> <i>Cittasvaram</i> 1-4

**Table 2.** The structure of the performance of *varnam* “Karunimpa” which is typical of *varnams* in carnatic music.

#### 5.2.1. Gamaka representation

In the transcription, *gamakā* fragments were represented using a simple numerical four-component representation -  $(p, a, s, r)$  - that can then be concatenated to form longer and more complex movements. The four components are as follows -

- $p$  = “focal pitch” expressed in semitones. A focal pitch is a quasi-stationary pitch point within a gamaka.<sup>4</sup>
- $a$  = attack duration - the time spent moving towards the focal pitch.
- $s$  = sustain duration - the time spent at the focal pitch.
- $r$  = release duration - the time spent moving away from the focal pitch.

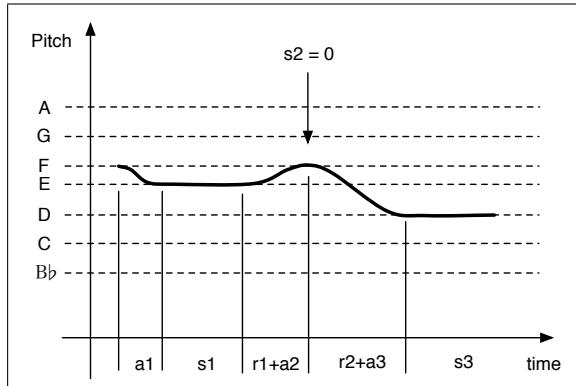
A full *gamakā* movement is built by concatenating such fragments using sinusoidal interpolation between their focal pitches over a time given by the sum of attack and release durations — i.e. in

$$(p_1, a_1, s_1, r_1) - (p_2, a_2, s_2, r_2),$$

the movement from  $p_1$  to  $p_2$  occurs over the period  $r_1 + a_2$ . Figure 1 illustrates the concatenation for the phrase *FEFD*. The redundancy between the attack and release durations is exploited to encode relationships of a gamaka fragment to the underlying pulse. Although linear interpolation of  $p$  is adequate for higher speeds, we've found that sinusoidal interpolation works better in slower speeds and becomes nearly linear in higher speed movements. Furthermore, sinusoidal interpolation helps preserve the quasi-stationariness of the focal pitches independent of speed.

In the rest of this paper, we refer to such *gamakā* fragments through their focal pitches and talk of focal pitches

<sup>4</sup>For ease of processing, the focal pitch is maintained as a pair of pitch values - the first giving the actual pitch and the second giving the intended pitch sans overshoots. The distinction can be ignored for the purpose of discussion.



**Figure 1.** Concatenating gamaka fragments *FEF* and *EFD* of phrase *FEFD* fuses their “attack” and “release” intervals using sinusoidal interpolation.

as “having” attack, sustain and release durations. Therefore, the *gamakā* of a “pluck” or “note” is talked about as consisting of a sequence of focal pitches.

It is worth noting that our approach bears some resemblance to the F0 contour models mentioned in §3. We interpolate in the  $\log(f)$  domain as in the Fujisaki model. The “maximum-based description” approach of Porte and Heuft uses sinusoidal interpolation, though in the  $f$  domain. The “second order linear system response” approach of the Fujisaki model also results in sinusoidal shapes when used with reduced damping.

### 5.2.2. Problems with pitch trackers

Automated pitch trackers were not used in this work due to the extensive normalization that needed to be done in the transcription to eliminate factors that are not relevant to the musical question at hand. The global tempo in the recording fluctuates around 71 beats per minute, but a fixed global tempo is adequate for this problem since compositions are specified that way. Metricity of gamakas are also bent locally for expressiveness. It is, again, adequate for our purpose to consider the closest strict-time equivalents of such phrases. Such a normalization requires familiarity with the genre and no known algorithms exist for this purpose. The tuning characteristics of the recorded instrument also change as it responds to environmental conditions such as heat and humidity. For the purpose of our inquiry, it is adequate to consider the pitches of the scale normalized to the equal tempered tuning system, removing the influence of such environmental conditions. Furthermore, pitch trackers, when they work, yield data in a form that is excessive for the purpose of understanding the music at the level of the problem at hand. If they were used, data reduction and verification of adequacy would be needed as well. With 592 plucks (notes) in the transcription of the whole composition featuring 1633 focal pitches between them, we felt that manual transcription served our purpose better than pitch trackers.

Trait	Speed	Value(s)
Prescriptive notes	Both	296
Plucks	1x	189
	2x	100
Focal pitches	1x	626
	2x	303
Unique pitch triads	1x	56
	2x	43
Note duration	1x	424-1697 ms, median = 848 ms
	2x	212-1060 ms, median = 212 ms
Gamaka duration	1x	53-1697 ms, median = 178 ms
	2x	25-848 ms, median = 107 ms

**Table 3.** Transcription statistics for the section of the analyzed performance that occurs in two speeds.

### 5.3. Melodic concordance comparison

To establish that the second speed rendition can, in principle, be derived from the first speed rendition, we needed to verify that the phrasings found in the second speed can all be accounted for in the first speed. We generated a “melodic concordance”<sup>5</sup> for the renditions in each of the speeds by extracting focal pitch n-grams that occur in sequence and forming a histogram of them. This gives us both the context in which each pitch occurs and how frequently it occurs in such a context. Digrams and trigrams were used and longer n-grams aren’t needed for this problem. Timing information cannot be used in this comparison because the problem under consideration is to calculate new gamaka timings upon change of speed.

In comparing melodic concordances, considering statistical significance alone is not enough. The exceptions also need to be studied since they might be musically significant and will need to be accounted for by the transformation rules. We found that all digrams in the second speed rendition were present in the first speed rendition as well. Out of 43 trigrams in the second speed (Table 3), five were not found in the first speed set, but they were accounted for as a) a discontinuous pitch movement that was present in the prescriptive notation, b) an acceptable equivalent in the first speed in the same time position, c) by an “oscillatory continuity” condition ( §6.4) and d) as part of an extra continuity pluck added by the performer. The anomalous pitch trigrams were *EFG*, *FDG*, *GFA*, *BbGA* and *FBbA*, where *C* is the tonic. It is worth noting that these trigrams are fast patterns lasting less than 1/2sec.

<sup>5</sup>A “concordance” is a tool used in literary analysis to study an author or poet’s writing style. It is a dictionary of words found in a collection of the author’s works, presented along with the contexts in which they were found.

#### 5.4. Analysis and rule construction

The “first speed” rendition is speed doubled without further changes in order to study the kinds of changes that need to be made to the gamakas. This also helps study the gamaka modifications necessary to emphasize the pulsing observed in the original second speed performance. The *cittasvaram* section of the performance can be used as a reference for studying the pulsing since it is only performed in the higher speed. Doing so leaves the original second speed performance intact for verification of the generated result.

#### 5.5. Re-synthesis and verification

Pulse emphasizing and complexity reducing transformations are performed on the slow speed version to generate a speed doubled version and the result is compared with the original double speed rendition. The comparison is based on a) the focal pitches used in each *gamakā* and b) the number of *gamakā* fragments that appear for each note given in the prescriptive notation. It is not meaningful to compare the timing data since their equivalence is directly encoded in the transformation algorithm.

### 6. TRANSFORMATION RULES

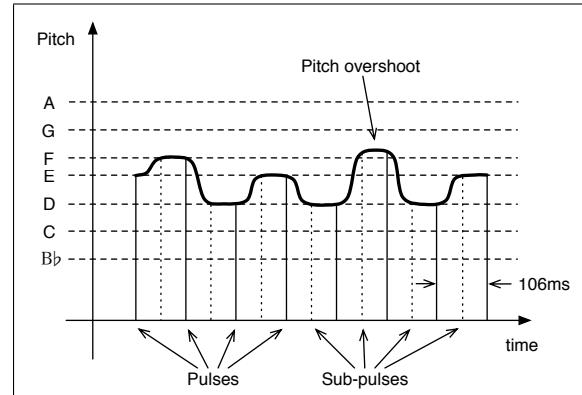
The following rules in conjunction with the prescriptive notation of the first part of the *varnam* (called *pallavi*, *anupallavi* and *muktayisvaram*, shown in Table 2) can be used to derive a musically close second speed rendition given a first speed rendition.

1. Speed limiting continuous pitch movements.
2. Aligning gamaka onsets to the underlying pulse.
3. Preserving focal pitches that have longer sustain times in the first speed phrases - relative to other focal pitches within the same phrase.
4. Dropping focal pitches based on whether they match the speeded up prescriptive data.

Tying together (2), (3) and (4) is an “oscillatory continuity” condition which further decides which focal pitches to add, preserve and drop in order to ensure that the gamakas of two adjacent notes can be joined using a “continuity pluck” [14]. These transformation rules were written as programs in the SuperCollider programming language widely used to build synthesizers. We now discuss these rules in detail.

#### 6.1. Movement speed limit

In the performance studied, the speed of continuous movement between two pitches had an upper limit of about 100ms per tone. String pulls and fret slides were treated in the same way since there is no such distinction in the vocal tradition of the genre which forms the basis. Movements occurring in the second speed hover around this “speed



**Figure 2.** Alignment of movement onsets to pulses and landing points to sub-pulses in the *gamakā* EFDEDDE.

limit” and therefore display a constant speed effect where more time is taken for deeper movements than for shallower movements. Pitch intervals larger than a tone take proportionately longer to cover. The focal pitch preservation and dropping rules come into effect when this speed limit is reached for a movement in the first step of simple speed doubling.

#### 6.2. Onset alignment of gamakas

The movement between two pitches were found to follow two types of pulse alignment in the slower speed - a) the onset of the movement aligns with a pulse and b) the landing point of the movement aligns with a pulse. The former dominated quicker intra-note movements and the latter occurred in slow fret slides.

In the second speed rendition, the dominant alignment is of the first kind. Therefore the transformer directly uses this information and aligns the onset of all gamakas on 1/8 count boundaries. To be precise, the onset of each gamaka fragment aligns with a 1/8 pulse and ends on the immediately following 1/16 pulse, as illustrated in Figure 2.

A special case occurs when two notes of durations 1 count and 2 counts occur in sequence in the first speed performance. The performer, on such occasions, may choose to symmetrize it by phrasing them both to be 1.5 counts long in the first speed. Such phrases were realigned to the 1+2 pattern before transforming for the second speed.

#### 6.3. Focal pitch preservation and dropping

For the purpose of this section, we view a *gamakā* as consisting of a sequence of focal pitches - for example FEFDF. *Gamakā* complexity is reduced by dropping certain focal pitches of a phrase while preserving others. The following rules were found to be adequate for this purpose. A “pre-processing” step for these rules is the removal of extra plucks in the slower speed. A pluck is considered extra if it features in the middle of a syllable of the lyrics. Extra plucks are inserted by *vīna* artists for audibility of long notes since the sound of the vibrating string decays over time.

### 6.3.1. Pulse assignment

Assign each focal pitch to an integer number of pulses. The sustain part of a focal pitch is to begin on a 1/16 sub-pulse and end on a 1/8 pulse, except if the focal pitch occurs at the start of a pluck, in which case the sustain part also starts on a 1/8 pulse. Movement is to last for half a pulse, unless overridden by the “speed limit” rule for large intervals. If more time is available, distribute pulses to the focal pitches which have longer sustain times in the slow speed *gamakā*. If less time is available, apply one of the dropping rules and try again.

One way to understand this transformation is by analogy to text to speech synthesis systems which time stretch vowels while preserving consonants. Focal pitches with relatively long sustains (within a pluck) seem analogous to vowels.

### 6.3.2. Stressed focal pitches

Preserve the ending focal pitch of a pluck in the transformation. The rationale for this rule might be that the following pluck on this ending focal pitch stresses it.

### 6.3.3. Dropping focal pitches

1. The first focal pitch of a pluck in the slower speed is dropped in the double speed rendition if it is a moving focal pitch - i.e. if it has zero sustain.
2. The first focal pitch of a pluck in the slower speed is also dropped in the double speed rendition if it has the same pitch value as the ending focal pitch of the preceding pluck. This pluck is then a “continuity pluck”. Note that this rule applies even if the starting focal pitch has a non-zero sustain duration.
3. If a prescribed pitch is assigned two focal pitches in the slow speed rendition and the time scaled movement is too fast in 2x speed, then the two focal pitches can be replaced with an stationary focal pitch (attack = release = 0) that is the same as the prescribed pitch.
4. An oscillatory pattern *xyxyxy* can be reduced to *xyxy* in the double speed version if not enough pulses are available to accommodate all the focal pitches and if it occurs in the middle of a *gamaka*.

## 6.4. Oscillatory continuity

When two successive notes in the second speed are such that at least one of them features an oscillatory *gamaka* and the adjacent note also has a movement, then additional movements continuing from the oscillation are added to the adjacent note in the second speed rendition, creating a feeling of continuity between them.

For example, the connected movement *DEDEF* in the slower speed, where the *DED* is of the same duration as the *E* and *F*, is transformed into *DEDfef* where the extra oscillation *DFE* has been added.

## 6.5. Microtonal adjustments

In addition to the above rules, microtonal adjustments to the focal pitch values of some movements are necessary for perceptual reasons - i.e. without an overshoot, the focal pitch sounds flatter than it actually is. This observation is consistent with vibrato studies which indicate that the perceived frequency of a note with vibrato is an average of the extreme frequencies [12, 7]. The occurrence of such overshoots in Carnatic music has been studied by Subramanian [16] and Krishnaswamy [9]. Subramanian also suggests that the intended pitch be approximated by a sliding window average. Figure 2 also illustrates one such overshoot occurring on the second *F* of the *gamakā EFDEDFDE* which occurs in the middle of the deep oscillation *DFD*.

Apart from perception, another reason for such overshoots could be the difficulty of precisely reaching pitches in fast oscillatory phrases using string pulling on the *vīna*. These two factors didn’t need to be separated for our purpose because we found that these overshoots are perceptually resilient to small variations ( $\sim \pm 10\%$ ) when evaluated in the context of a phrases that are several seconds long. Therefore there is no reason to suspect that the effect of the skill dependent physical precision constraint is significant for the purpose of resynthesis.

These findings were incorporated into the following rules -

1. Only overshoots occur, no “undershoots”. It is likely that this is a consequence of the use of the *vīna* in the performance. The *vīna* being a fretted stringed instrument, it is only possible to increase the pitch by pulling on the string from a particular fret. In other performance modes such as singing or violin playing, undershoots could occur.
2. Only focal pitches with sustains of 1/16 of a count - i.e. of the duration of a sub-pulse - are given non-zero overshoots. Those with sustains of 1/8 or longer are not given given any overshoots.
3. A “depth” is assigned to an oscillation of the form *xyz*, where *y* is the highest pitch of the three, that is equal to one less than the number of semitones of the smaller of the two intervals *xy* and *yz*.<sup>6</sup> For all other types of *xyz* movements, the depth of *y* is set to zero.

$$\text{depth}(xyz) = \max(0, \min(3, y - x, y - z) - 1) \quad (1)$$

4. Applied overshoot =  $\text{depth} \times 25$  cents.

The above rules were adequate for most of the overshoots found. One phrase in the slower speed was transcribed with an overshoot of 80 cents and we acknowledge that there is an unavoidable ambiguity in this case. The phrase is *GAGAG* and its execution is closer to *GBbGBbG*.

<sup>6</sup>Due to the way we’ve defined “focal pitch”, two consecutive focal pitches within a single *gamaka* cannot be the same.

This deep overshoot, however, disappears in the double speed rendition where our depth rule matches the performance. The strangeness of the slower speed rendition could be because the performer spends more time on the first and last *Gs* in the phrase, causing the movements in the middle to be, ironically, faster than in the pulse aligned double speed rendition. Though this suggests that the overshoot depends on the slope, we didn't need to account for it since the above interval rule was adequate to generate a comparable double speed performance.

## 7. DISCUSSION

The reduction in detail is a necessary component of a change of speed, with the degree of performed detail approaching that given in the prescriptive notation in even faster speeds. The higher speed renditions reflect a performer's skill and taste as well the general listener's ability to appreciate the performer's choices. A straightforward speed up of the slow speed performance yields *gamakās* that are melodically valid but would sound very strange to someone familiar with the genre due to the difficulty of physically performing them. There is considerable detail in the timing of the movements in the slow speed that sometimes translate to "inhuman" precision in timing in a straightforward speed up. Many movements that occur over a single sub-pulse (1/16 of a count of the 16 count *tālā* cycle) would become too fast to perform over a duration of 1/32 of a count in the straightforward speed up version. Therefore imposing a speed limit on movements causes any timing difference between two consecutive movements in the slower speed rendition to get ironed out in the higher speed rendition. This effect is similar to what has been noted in [4] and [6] about the tempo dependence of jazz swing ratios, where the swing ratio was seen to approach 1:1 as the tempo increased.

The rules found in this study are not generalizable across ragas and performance styles because it is based on a single performance of a composition. However, the forms of the rules appear to be generalizable to other ragas, at least within the same performer's style. This could be because the raga and performance idiosyncrasies have been isolated to the original first speed performance. For example, the pitch overshoot rule described in §6.5 and the focal pitch manipulation rules of §6.3 are *rāgā* independent in form. However, the oscillatory continuity rule of §6.4 is a subtle one and, although *rāgā* independent in implementation, could potentially have incompatibilities with other styles and *rāgās*.

The rules presented here yield parameters for controlling the synthesis output. Though the values of the parameters used in this study are derived from the original first speed performance, changing them can yield different and potentially interesting re-synthesis results. For example, the "speed limit" rule's speed parameter controls the complexity that can feature in the second speed. Decreasing this parameter will decrease the complexity and increasing it will eventually preserve all the details in the

first speed performance, save for pulse alignment. Similarly, choosing a 1/6 pulsing instead of 1/8 creates a very different "broken tisram" feel to the second speed. Although such a feel would not be accepted by convention, local shifts to tisram are used by performers for variety and this parameter can help introduce such variety into the re-synthesis.

We now discuss the broader context of our work and directions for extension.

## 8. FUTURE WORK

The shape of *gamakās* that our simple sinusoidal interpolation approach generates, although it is sufficient for the purpose of our study, covers limited ground. In particular, allowing for asymmetry in the shape of the movement would yield a noticeable improvement in re-synthesis fidelity, though at the cost of transcription complexity. Physical aspects of the instrument and performer also influence *gamakā* shapes. Carnatic music is a vocal tradition and therefore there is a tendency for instrumental performance to mimic singing. Musicians also cross-train - i.e. a musician focusing on a particular instrument might choose to study under another musician who specializes on a different instrument. This cultural aspect suggests that the physical aspects of multiple instruments, including the voice, might have a part to play in shaping the movements executed by a performer. *Gamakā* shapes also depend on instrumental technique and a performer may have to choose from among a few techniques to execute an abstract *gamakā*. The choice of technique at such points might partially reflect the performer's physical situation with the instrument. At the semantic level, song lyrics suggesting physical movements such as swinging might inspire a performer to suggest them with appropriately shaped *gamakās*. Extending our shape model to cover more of this diversity is therefore an interesting direction.

Generalizations of our work along the axes of *rāgā* and performer style require similar intensive study of performances in varied *rāgās* by artists belonging to different schools of training. In particular, the *rāgā* axis offers great variety and challenge for such study. Given the intensive nature of such a study, the broader problem of modeling *rāgā* and the performer's style as a synthesizer for prescriptive notation would be a more fruitful though challenging direction to take. Several factors relating to performance appear in a unified form in such a synthesizer including influence of the physics of the instrument's controls on the specifics of the rendition, the style(s) that the performer schooled in, the performer's own idiosyncrasies and the broad genre characteristics part of the standard pedagogy. Parameterizing these factors in such a synthesizer can yield valuable tools for teaching and study of carnatic music and its styles.

## 9. CONCLUSION

For one *rāgā*, *tālā* and performer style, strengthening pulse relationships and speed limiting *gamakās* in conjunction with *gamakā* simplification rules can be used to model the detail reduction seen in two-speed performances of *varnams* in carnatic music. Microtonal adjustments of the focal pitch of *gamakas* are important to generate the requisite rhythms in the higher speed. The raga-independent nature of some of the constructed rules suggest that they may characterize the style of the performer and therefore could be transportable to other ragas in the performer's style.

## 10. REFERENCES

- [1] G. Berndtsson, "The KTH rule system for singing synthesis," *Computer Music Journal*, vol. 20, no. 1, p. 76–91, 1996.
- [2] G. L. Collier and J. L. Collier, "The swing rhythm in jazz," in *Proceedings of the 4th International Conference on Music Perception and Cognition*, 1996, p. 477–480.
- [3] A. Friberg, V. Colombo, L. Frydén, and J. Sundberg, "Generating musical performances with director musices," *Computer Music Journal*, vol. 24, no. 3, p. 23–29, 2000.
- [4] A. Friberg and A. Sundström, "Swing ratios and ensemble timing in jazz performance: Evidence for a common rhythmic pattern," *Music Perception*, vol. 19, no. 3, p. 333–349, 2002.
- [5] H. Fujisaki, "Dynamic characteristics of voice fundamental frequency in speech and singing. acoustical analysis and physiological interpretations," in *Proceedings of the 4th FASE Symposium on Acoustics and Speech*, vol. 2, 1981, p. 57–70.
- [6] H. Honing and W. B. de Haas, "Swing once more: Relating timing and tempo in expert jazz drumming," *Music perception*, vol. 25, no. 5, p. 471–476, 2008.
- [7] Y. Horii, "Acoustic analysis of vocal vibrato: A theoretical interpretation of data," *Journal of voice*, vol. 3, no. 1, p. 36–43, 1989.
- [8] A. Kirke and E. R. Miranda, "A survey of computer systems for expressive music performance," *ACM Computing Surveys (CSUR)*, vol. 42, p. 3:1–3:41, Dec. 2009, ACM ID: 1592454.
- [9] A. Krishnaswamy, "Application of pitch tracking to south indian classical music," in *Proc IEEE ICASSP*, 2003.
- [10] A. Monaghan, "State-of-the-Art summary of european synthetic prosody R&D," in *Improvements in speech synthesis: COST 258: the naturalness of synthetic speech*, E. Keller, G. Bailly, M. A. J. Terken, and M. Huckvale, Eds. Wiley, 2002, pp. 93–103.
- [11] T. Portele and B. Heuft, "The maximum-based description of f0 contours and its application to english," in *Fifth International Conference on Spoken Language Processing*, 1998.
- [12] E. Prame, "Measurements of the vibrato rate of ten singers," *Journal of the Acoustical Society of America*, vol. 96, no. 4, p. 1979–1984, 1994.
- [13] C. Seeger, "Prescriptive and descriptive Music-Writing," *The Musical Quarterly*, vol. 44, no. 2, pp. 184–195, Apr. 1958.
- [14] K. S. Subramanian, "South indian vina tradition and individual style," Ph.D. Thesis, Wesleyan University, Connecticut, USA, 1985.
- [15] M. Subramanian, "GAAYAKA - carnatic music notation player." [Online]. Available: <http://carnatic2000.tripod.com/gaayaka6.htm>
- [16] —, "Analysis of gamakams of carnatic music using the computer," *Sangeet Natak*, vol. XXXVII, no. 1, pp. 26–47, 2002.
- [17] —, "Carnatic music - automatic computer synthesis of gamakams," *Sangeet Natak*, vol. XLIII, no. 3, 2009.
- [18] J. Sundberg, A. Askenfelt, and L. Fryden, "Musical performance: A Synthesis-by-Rule approach," *Computer Music Journal*, vol. 7, no. 1, pp. 37–43, 1983.
- [19] P. Taylor, "The rise/fall/connection model of intonation," *Speech Communication*, vol. 15, no. 1-2, p. 169–186, 1994.
- [20] —, "The tilt intonation model," in *Fifth International Conference on Spoken Language Processing*, 1998.
- [21] P. Taylor and A. W. Black, "Synthesizing conversational intonation from a linguistically rich input," in *The Second ESCA/IEEE Workshop on Speech Synthesis*, 1994.
- [22] T. Yoshimura, K. Tokuda, T. Masuko, T. Kobayashi, and T. Kitamura, "Duration modeling for HMM-based speech synthesis," in *Fifth International Conference on Spoken Language Processing*, 1998.